2-20-2014

SAGNAS (Group 9) Project Proposal and Plan

- Plan: Feb. 21 (F) This Deadline
- First Increment: Mar. 7 (F)
- Second Increment: Mar. 21 (F)
- Third Increment: Apr. 11 (F)
- Fourth Increment: Apr. 25 (F)
- Project Presentations: May 6 (T)& May 8 (Th)
- Final Project: May 9 (F)

I. Introduction

Our motivation for this application is to give users from major Cities like Chicago, Kansas City, New York, Seattle, etc. in The United States the ability to be more knowledgeable about their surroundings by providing them with crime updates and predicting how safe they are at any point of time based on their current location. We also provide the ability to share one's geo-location with other users of the application via shared rooms.

II. Project Goal and Objectives

Our objectives are to provide accurate, easy to use information about local crime events for anybody to use. In doing so, we hope to educate people who want to take every measure to be as safe as possible. By providing a prediction system based on the crime history based on a specific region. Additionally, we want to let people feel safer by giving them the opportunity to share their location with people they trust at any time, regardless of Wi-Fi connection (GPS only is necessary).

The system will have an additional functionality so that users will no longer have to worry when travelling alone. This feature shares your location information with your partners/friends/anyone who they trust and notify them using a single click of their device.

We haven't found an application that provides the combination of crime event display along with predictive recommendations based off that information on the mobile platform, and we

III. Project Background and Related Work

- Google Hangout (Real-time connection and live updates through single interface)
- Livestream (User broadcasts information to others. In our case, it's a two way broadcast)
- GPS Tracking (running apps, can cite specific one e.g. Runkeeper)
- Real-time position sharing (Google Maps)

IV. Proposed System

1) Requirement Specification

Functional Requirements

- Users create group rooms, where they can share their geolocation and view other room members' geolocations
- Geolocation display/sharing management is done through the Multiway Coordination web service
- The group's geolocation data will be shown on a map, which will additionally have crime rates overlaid on top
- Crime event datasets will be stored in HBase
- Before using Mahout to provide user recommendations, datasets must be exported from HBase
- Crime event data will be analyzed using Mahout's machine learning algorithms, to notify users of relevant trends and rates of crime as they travel around the map
- Mahout's outputs will be stored in HBase as a separate table
- Solr will index all HBase data for easier, faster access
- Web services will display relevant information accessed from Solr onto the mobile map interface
- Crime statistics based on the crime event data will be accessed and displayed via a separate interface for users

Non-functional requirements

- Geolocation display/sharing should occur with minimal delay
- Crime events should be shown in an easy to understand manner
- Any relevant crime patterns or trends should be displayed simply, so the user can make appropriate decisions

Technical and Business requirements

- We will use Amazon AWS services for cloud storage
- Cloudera's tools will be used for all Hadoop-related functionality
- For project management, we must use ScrumDo
- Github must be used to version control
- The following APIs must be used:
    - Google Maps

Business Process/Workflow analysis

- Determine the highest priority tasks for current increment
- Work on tasks
- Review finished tasks at end of the week
- Keep communication lines open throughout the week as questions/comments from each team member arise

Technological and Architectural requirements

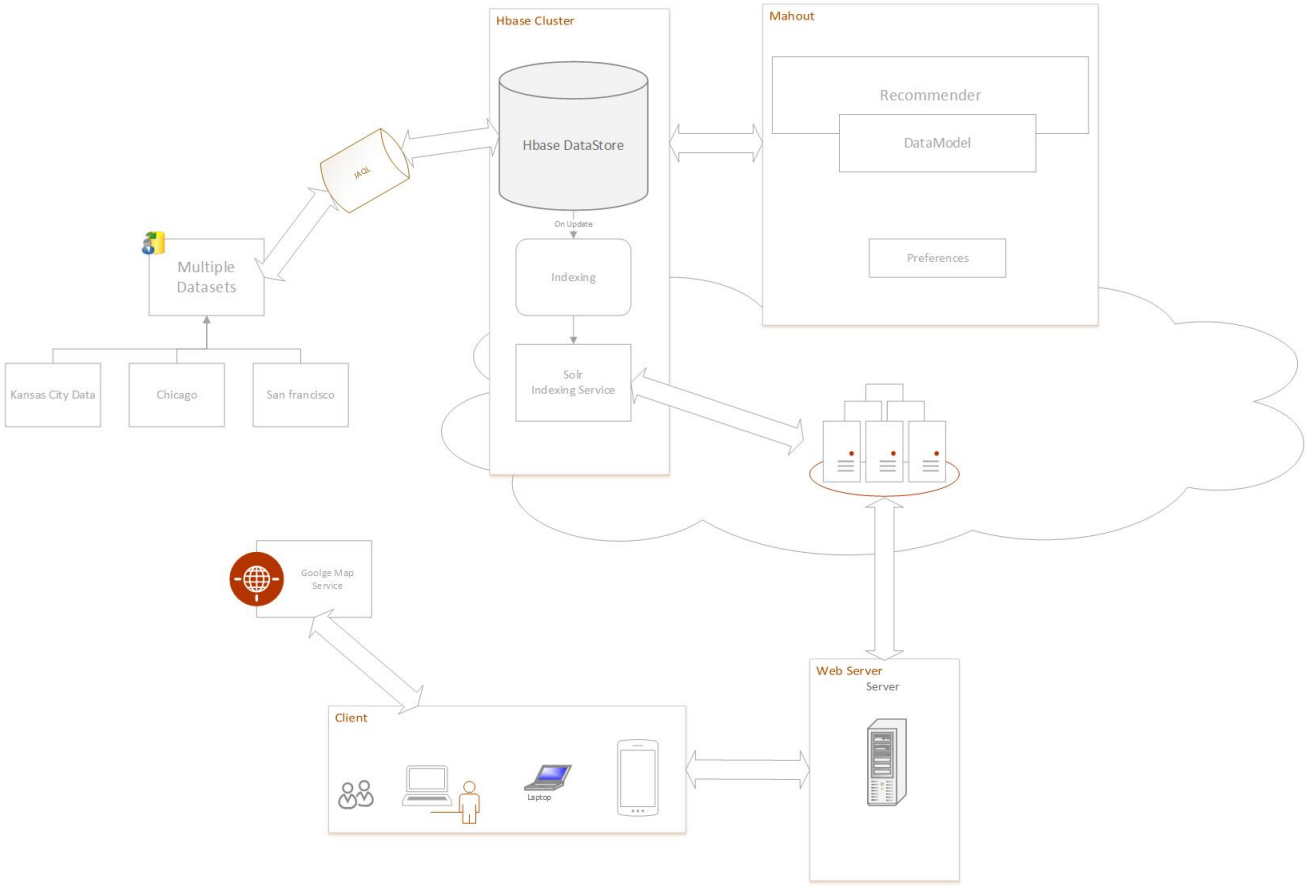
- Cloudera Virtual Machine and Hadoop tools
- Hadoop
- HBase
- Mahout
- Solr
- Solr's RESTful services
- Google Maps API
- Compatibility with Firefox and Google Chrome
- Twitter Bootstrap

2) Framework Specification:

- Users access the application via a web browser
- Internet connection is necessary for crime display/recommendations
- Two different types of datasets will be stored in HBase:
    - The crime event data set, which be used to display events on a map, as well as general crime statistics
    - The Mahout generated user recommendations, which, based on zip code, will provide users with notifications if their current travelling area has a high crime rate for a specified type of crime
- Methodologies and Algorithms
    - We will be using machine learning algorithms within Mahout, but have yet to determine which algorithm will be necessary

System Architecture Diagram

System Architecture of SAGNAS



The client interacts with the SAGNAS service via a web server, and Google Maps via an API. The SAGNAS service is stored in the cloud. Web services will interact directly with Solr, which searches/indexes all relevant data from HBase. HBase stores crime event datasets, which are parsed and restructured for appropriate usage via JAQL, and user recommendation datasets, which are generated via Mahout's machine learning functionalities.

3) System Specification:

- Datasets

We will be utilizing three crime event datasets, from three cities: Kansas City, San Francisco, and Seattle. along with lat/long coordinates and crime description.
KC:https://data.kcmo.org/Crime/KCPD-Crime-Data-2014/yu5f-iqbp.json

SF:https://data.sfgov.org/Public-Safety/SFPD-Reported-Incidents-2003-to-Present/dyj4-n68b

Seattle:https://data.seattle.gov/Public-Safety/Seattle-Police-Department-Police-Report-Incident/7ais-f98f

New Services to be built:

- User recommendation service
  - Recommendations will be calculated weekly (not real-time)
  - Calculations will take the crime event dataset as an input, and produce a user recommendation dataset as an output
  - Will notify users of rate of crime happening in a certain area
  - Will utilize Mahout generated recommendation dataset accessed from HBase via Solr

- Crime information service
  - The input it uses is the crime event dataset
  - Will be displayed using charts and tables, using information based on the crime event dataset stored in HBase, indexed in Solr
  -

- Crime display service
  - Will be displayed on Google Maps, using information based on the crime event dataset stored in HBase, indexed in Solr

- Design of Mobile Client
  - We will utilize HTML5, Twitter Bootstrap, AJAX and CSS3

V. Plan by Services (using ScrumDo)

- ScrumDo URL: https://www.scrumdo.com/projects/project/kdm-sagnas/

VI. Bibliography

- http://www.chrome.com/campaigns/rollit/
- http://new.livestream.com/
- https://developers.google.com/accounts/
- https://developers.google.com/maps/